Inferential rules for confirmatory robustness

Aki Lehtinen

P.O. Box 24 00014 University of Helsinki, Finland

aki.lehtinen@helsinki.fi

tel. +358407002044

Paper forthcoming in European Journal for the Philosophy of Science. This is not the last version.

Abstract

This paper explores the conditions under which the robustness of a result provides confirmation, specifying what is confirmed and the degree of confirmation it offers. Two inferential rules are proposed for modellers and experimenters to assess the impact of robustness on confirmation. These rules apply to both derivational and experimental robustness, but they are insufficient for empirical confirmation from derivational robustness, which requires the right kind of indirect confirmation relations. Several such relations are analysed in detail. When derivational robustness leads to empirical confirmation, it does so by increasing indirect confirmation, demonstrating that a model result (R_M) and an empirically validated result (R) depend on the same confirmed assumptions by showing the irrelevance of certain false auxiliary assumptions to the model result.

keywords: Derivational robustness; Experimental robustness, Indirect confirmation; Genuine confirmation; Logical omniscience

1 Introduction

A scientific result is considered robust if it is detected by several diverse means. More specifically, a model result, R, is said to be *derivationally* robust if it can be derived from several models that share a core structure, C, and vary in their auxiliary assumptions A_i.[1] Proponents of robustness (e.g., Weisberg 2006; Kuorikoski et al. 2010; Lloyd 2015) argue that if the core structure C plays a role in every demonstration of the robustness of R, this increases modellers' confidence in the *robust theorem*: 'ceteris paribus, if C then R'.

A number of recent contributions have made the stronger claim that robustness confirms (Lloyd 2015; Schupbach 2015, 2018; Lehtinen 2016, 2018; Winsberg 2021; Dethier 2024; Casini & Landes 2024). However, given that each of these accounts differs, it is evident that more work remains to be done.

There is a substantial body of criticism concerning the supposed confirmatory benefits of robustness.[2] Given this state of affairs, my aim is to develop an account of confirmatory robustness that can address at least the most central of these criticisms.

Instead of reviewing the objections here, let me offer a clarification. The formulation of the robust theorem may suggest that R is highly probable given C. A confirmation theorist might interpret the theorem as implicitly invoking absolute confirmation—that is, as claiming that the conditional probability p(R|C) exceeds some threshold. While many theorists are satisfied with the conventional view that absolute confirmation requires only p(H|E)>1/2, one might reasonably insist, especially for the theorem's applicability to real-world cases (see Harris 2021; Harris & Frigg 2023), that p(R|C) must be substantially higher. While many of the criticisms are apt

[1] I use the terms derivational, measurement, and inferential robustness in roughly the sense outlined by Woodward (2006), with the following exceptions. Given the convergence in the literature on the requirement that a model must possess a core structure, I take derivational robustness to imply such a commitment, even though Woodward did not. I also use experimental and measurement robustness interchangeably, whereas Woodward did not employ the former term.

[2] Without attempting to be exhaustive, the most prominent challenges have been directed at derivational robustness (Orzack and Sober 1993; Forber 2010; Calcott 2011; Odenbaugh and Alexandrova 2011; Houkes and Vaesen 2012; Woodward 2006; Parker 2011; Lisciandra 2017; Harris 2021, Mcloone et al. 2025). See Stegenga & Menon (2017), or Hudson (2013) for experimental robustness.

if robustness were claimed to provide absolute confirmation. However, it only provides incremental confirmation, and the arguments for and against it can be framed in terms of arguments for and against the inferential rules proposed here.

This paper has two main aims: first, to formulate inferential rules that explain and justify an increase in p(R|C) when R is shown to be robust, thereby meeting the critics' challenge to accounts of confirmation by robustness; and second, to identify the kinds of confirmation produced by such increases in different contexts. I argue for the rationality of adopting two *inferential rules* that justify an increase in p(R|C). These inferential rules guide how modellers and experimenters should adjust their subjective beliefs concerning p(R|C) in light of new derivational or experimental information regarding the components involved in deriving robust results.

The inferential rules developed here respond to a recent challenge by McLoone, Orzack, and Sober (2025), who argue that existing accounts of confirmation by robustness rely on conditioning on tautologies and therefore fail to show how robustness yields epistemic gain. Their objection, however, depends on standard probabilistic assumptions—most notably the idealisation of agents as logically omniscient—and they explicitly invite non-standard probabilistic approaches. The rules introduced below are intended in this spirit: they form a non-standard probabilistic framework applicable only when agents are not logically omniscient, within which non-empirical confirmation by robustness gains clear epistemic significance.

Broadly speaking, the first rule states that employing a component X in a model or experiment justifiably increases the conditional probability p(R|X), provided that suitable variation is introduced in the other components. The second rule holds that deriving or observing a result R without component X decreases p(R|X). In modelling contexts, p(R|C) increases through robustness when certain auxiliary assumptions—initially thought to be required for deriving R—are later shown to be dispensable. Demonstrating that R is not logically dependent on these auxiliaries strengthens the link between R and the core model element C, thereby increasing p(R|C).

The kind of confirmation achieved when p(R|C) increases through robustness differs between modelling and experimentation. In modelling, as Orzack and Sober (1993) note, robustness does not affect the confirmation of R by direct empirical evidence E: since p(E|R) remains unchanged, R cannot be more confirmed by E merely because it is robust. Thus, an increase in p(R|C) yields only non-empirical confirmation, and it is essential to distinguish this

from empirical confirmation when assessing derivational robustness.

In contrast, in experimental contexts R itself constitutes empirical evidence, and experimental robustness contributes directly to empirical confirmation in a way that derivational robustness cannot without empirical support. More generally, although it is desirable to formulate accounts applicable across types of robustness for the sake of generality, the contextual differences between modelling and experimentation are substantial enough to require distinct accounts for derivational and measurement (or experimental) robustness.

The inferential rules are formulated for both derivational and measurement robustness, though they differ in application. Because applying these rules in modelling yields only non-empirical confirmation, they must be supplemented by an account of how empirical evidence bears on the model family's components and results. By applying the rules to the evidential relations identified by Lloyd (2015) and Lehtinen (2016, 2018), I simplify, reformulate, and extend Lehtinen's account of indirect confirmation. Lloyd's treatment is conceptually vague, and my earlier presentation unclear, which may explain why the account has received little attention (but see McLoone 2025). This paper aims to clarify and make these ideas more accessible.

Lehtinen's account warrants further development for two reasons. First, if Orzack and Sober are right that derivational robustness cannot yield direct empirical confirmation, then indirect confirmation becomes indispensable for linking robustness to empirical evidence. Second, the version developed here subsumes the cases discussed by Lloyd (2010, 2015) and Casini and Landes (2024). Finally, although the inferential rules are not framed as independence conditions, they capture the same kind of weak independence as Schupbach's (2018) Robustness Analysis independence, thereby avoiding the objections that threaten stronger independence assumptions (Schupbach 2015, 2018; Harris 2021).

Although the inferential rules and their preconditions draw on aspects of Schupbach's account, they do not collapse into his notion of robustness as explanatory discrimination. The rules are expressed in terms of changes in the expectedness of results rather than explanatory discrimination. One might object that this distinction is merely verbal, since Schupbach models explanatory power probabilistically (Schupbach & Sprenger 2011), thereby treating expectedness and explanatoriness as equivalent.

However, recasting Schupbach's account explicitly in terms of expectedness exposes a problem. His formal treatment assumes that a new detection

of the robust result will occur with a probability near one if the target hypothesis explains it. This is plausible only when the models or experiments in the known and anticipated detections are so compositionally similar that the difference is trivial—that is, when the changes in auxiliary assumptions are already expected to be irrelevant. By contrast, the inferential rules presuppose genuine uncertainty about at least some auxiliaries. Near-certainty in expectations about future detections is possible but neither assumed nor required.

In principle, the account developed here applies across disciplines. The Lotka–Volterra model serves to illustrate the inferential rules in Section 2, and climate modelling demonstrates their use in empirically informed contexts in Section 4, though no detailed case study is included for reasons of length. The paper proceeds as follows: Section 2.1 introduces the inferential rules and provides an intuitive overview of their operation; Sections 2.2 and 2.3 defend them—2.2 by relating them to genuine confirmation and 2.3 by explaining their preconditions and differences, despite certain similarities, from Schupbach's account; and Section 3 offers a comprehensive account of the empirical confirmation relations arising from derivational robustness.

2 Inferential rules for changing conditional probabilities

2.1 Notation, and an intuitive introduction with the Lotka-Volterra model

This section begins with a well-known example of robustness—the Volterra principle—which illustrates the kinds of circumstances to which the inferential rules apply. Only a minimal presentation of the Lotka–Volterra model is given, since the case has been extensively discussed by philosophers (e.g., Weisberg & Reisman 2009; Knuuttila & Loettgers 2017; Schupbach 2018; Harris 2021). I then introduce the inferential rules and show how they operate in this example. With this intuitive case in place, I proceed to examine the conditions for applying the rules, the arguments for their rationality, and the presuppositions underlying their use.

The Lotka–Volterra model describes the population dynamics of a two-species predator–prey system. It consists of two differential equations

$$\begin{cases} \dot{x} = \alpha x - \beta xy \\ \dot{y} = -\gamma y + \delta xy, \end{cases} \tag{1}$$

where x denotes prey, y denotes predator, αx is natural prey growth, −βx prey lost to predation, −γy natural predator mortality, and δxy predator growth from consuming prey. The Volterra principle states that a proportional biocide increases prey abundance and decreases predator abundance when the system is negatively coupled—that is, when each species' growth rate depends on the other's abundance with opposite signs ($\partial\dot{x}/\partial y < 0$,and $\partial\dot{y}/\partial x > 0$). Let C denote 'the predator–prey system is negatively coupled' and R denote 'a general biocide increases prey abundance and decreases predator abundance'. Lotka and Volterra first derived R under a set of assumptions represented schematically as:

$$M_1 = (CA_1A_2A_3) \vdash R \tag{2}$$

Here, A_1 assumes that prey cannot take cover, A_2 that predator growth is density-independent, and A_3 that a biocide affects predators and prey proportionally. Modellers later showed that the Lotka–Volterra model is robust with respect to A_1 by replacing it with $A_1{}'$, which specifies that prey can take cover:

$$\begin{aligned} M_1 &= (CA_1A_2A_3) \vdash R \\ M_2 &= (CA_1'A_2A_3) \vdash R \end{aligned} \tag{3}$$

Given these derivations, how should the modeller revise her belief about p(R|C)? If the inferential rules are accepted, their application justifies an increase in this conditional probability. Probabilities such as p(R|C) and p(R|A_1) depend on background knowledge B, which includes information about the derivational relations among model components within a model family. Let p(R|C,B_0) denote the probability of R given C before the modeller knows that R is robust (the epistemic situation B_0 in (2)), and p(R|C,B_1) the corresponding probability after robustness has been established (in (3)).

In modelling, establishing that p(R|C,B_1)>p(R|C,B_0) constitutes *non-empirical confirmation*, since the conditional probabilities change with beliefs about derivability. This is a case of *logical learning* and thus departs from

the assumption of *logical omniscience* (Garber 1983). The prior probability $p(R|C,B_0)$ reflects epistemic uncertainty about derivability relations, which is reduced as modellers update their beliefs to $p(R|C,B_1)$.

In standard Bayesianism, the prior $p(H|B)$ and conditional probabilities $p(E|H,B)$ are defined on the assumption that background knowledge B is fixed and logically closed when new evidence E is received. The inferential rules address the converse problem: how background knowledge changes when new derivational information arrives, while knowledge of empirical evidence remains fixed. In other words, the aim is to study how $p(R|C,B)$ and $p(R|A_i,B)$ vary with derivational information when $p(E|R)$ and $p(E)$ are held constant. Accordingly, expressions such as $p(R|C,B)$ and $p(R|A_i,B)$ should not be interpreted in the standard way. Under the usual interpretation, $p(R|A_i,B)$ would be undefined, since A_i is known to be false and thus $p(A_i|B)=0$ in the ratio $\frac{p(R\&A_i|B)}{p(A_i|B)}$.

The inferential rules describe how conditional probabilities $p(R|C)$ change as new derivational information from models becomes available. Under logical omniscience, such probabilities would be meaningless. The expression $p(CA_1A_2A_3)\vdash R)$ represents the modeller's belief about whether R can be derived from the conjunction $(CA_1A_2A_3)$ before the derivation is carried out. Ultimately, however, the focus is on the role of an individual component X in the derivation. The most natural interpretation of $p(R|X, B)$ is 'how likely is it that X is necessary for deriving R?'—denoted $p(X\vdash_C R|B)$ (The rationale for this notation will be provided in Sect. 2.2.). Such beliefs concern derivability or expectedness. When these beliefs are applied to evaluating how empirical evidence E (where $p(E|R)>p(E)$) confirms X or a result derivable with X, the background belief $p(X\vdash_C R|B)$ helps determine $p(R|X,B)$ together with other information. This conditional probability is interpreted almost standardly, except that B need not be logically closed; it reflects modellers' subjective beliefs about how E bears on X through the strength of $p(R|X,B)$.

The probability of being able to derive R from M_2, given background knowledge B_0: $(CA_1A_2A_3)\vdash R$ is expressed as $p((CA_1'A_2A_3)\vdash R|(CA_1A_2A_3)\vdash R$), where the solidus denotes conditioning on derivational information. This knowledge underlies the belief $p(R|C,B_0)=p(C\vdash_C R|B_0)=p(C\vdash_C R|CA_1A_2A_3)\vdash R)$, which represents the probabability that C is necessary for deriving R in a model family, given the background derivability information $(CA_1A_2A_3)\vdash R$. Although $(CA_1A_2A_3)\vdash R$ becomes a tautology once R is derived from M_1,

this does not make $C \vdash R$ tautological, since C alone typically does not entail R in cases of robustness. The expression p(R|C,B_1) abbreviates p($C \vdash_C R | (CA_1A_2A_3) \vdash R, (CA_1'A_2A_3) \vdash R$).

Strengthening the robust theorem occurs when p(R|C,B) increases as R is derived from multiple models that all include C, such that

$$p(R|C,B_1) > p(R|C,B_0) \leftrightarrow$$
$$p(C \vdash_C R | (CA_1A_2A_3) \vdash R, (CA_1'A_2A_3) \vdash R) > p(C \vdash_C R | CA_1A_2A_3) \vdash R).$$

Earlier we considered p($CA_1A_2A_3) \vdash R$), which in this framework can equivalently be expressed as p(R|$CA_1A_2A_3$). AlthoughA_1 is known to be false, this probability is not undefined, since conditioning here concerns derivability, not the truth of $CA_1A_2A_3$.

In what follows, p(X$\vdash_C$R|B) and p(R|C,B) are used interchangeably. Although it would be more economical to retain one notation, p(R|C,B) is kept to highlight how changes in p(X$\vdash_C$R|B) affect the evaluation of empirical evidence through p(R|C,B). Using p(R|C,B) alone would be misleading, since p(R|B) is undefined unless the basis of derivation is specified. To express that an element X is believed irrelevant to deriving R, it is therefore more natural to write p(X$\vdash_C$R|B)=0 rather than p(R|X,B) = p(R|B). Conversely, when a modeller believes X may be relevant or necessary for R, this is captured by p(X$\vdash_C$R|B)>0, which expresses the strength of that belief.

There are two inferential rules: one governing probability increase and the other decrease. The Derivational Confirmation Rule (DCR) and the Derivational Disconfirmation Rule (DDR) are defined as follows:

DCR: If a result R is derived using a conjunction of identified elements that include X, and this specific set of elements has not previously been used to derive R, then the conditional probability p(R|X) justifiably increases: p(R|X,B_1)>p(R|X,B_0).

DDR: If a result R is derived using a conjunction of identified elements that do not include X, then the conditional probability p(R|X,B_1) decreases, or the probability of derivability p(X$\vdash_C$R|B_1) is set to zero.

Let us now examine how applying the inferential rules justifies the inequality p(R|C,B_1)>p(R|C,B_0) in this example. Before Volterra derived the Volterra principle, it is plausible that his contemporaries had little confidence that negative coupling best explained the wartime increase in predator abundance

in the Adriatic Sea. Let B_{-1} denote this background knowledge prior to eq. (2). According to the Derivational Confirmation Rule DCR, $p(R|C,B_0) > p(R|C,B_{-1})$, since the derivation of R included C and had not been carried out previously. Although Volterra may have suspected that the auxiliary assumptions about prey cover (A_1) and predator density independence (A_2) were irrelevant to the relation betwen C and R, DCR likewise implies $p(R|A_1,B_0) > p(R|A_1,B_{-1})$, and analogously for A_2 and A_3.

When the Volterra principle is shown to be robust by deriving (3), applying DCR yields $p(R|C,B_1) > p(R|C,B_0)$, $p(R|A_2,B_1) > p(R|A_2,B_0)$, and $p(R|A_3,B_1) > p(R|A_3,B_0)$, since M_2 introduces a new set of elements from which R is derived with these components. In contrast, applying DDR to A_1 gives $p(R|A_1,B_1) < p(R|A_1,B_0)$, because R was derived without A_1. The inequalities for C and A_3 indicate that the Volterra principle is *non-empirically confirmed* by the derivation of (3). According to DCR, the same derivation also justifies $p(R|A_2,B_1) > p(R|A_2,B_0)$, though this does not imply that A_2 is necessary for deriving R - indeed later studies have shown that the principle is somewhat robust to changes in A_2.

Although DCR can apply to entirely new results, demonstrating the robustness of an existing result also requires DDR. Once a result has been derived at least once, p(R|C) increases only if some auxiliary that might have produced R is shown to be dispensable. In this case, the assumption that prey cannot take cover (A_1) was such an auxiliary. Applying DDR allows modellers to conclude that $p(R|A_1,B_0) > p(R|A_1,B_1)$ even though DCR previously implied $p(R|A_1,B_0) > p(R|A_1,B_{-1})$. The Volterra principle therefore does not depend on the assumption A_1. Indeed, once A_1 is shown to be irrelevant to R, DDR entails $p(A_1 \vdash_C R|B_1) = 0$.

Before examining the rationality of the inferential rules in detail, it is useful to outline their intuitive basis. Showing that some elements thought necessary for deriving a result are in fact dispensable rationally increases confidence that the remaining elements are essential. Even some critics of robustness implicitly accept this intuition and, by extension, DCR. For instance, Harris (2021) critiques the independence condition proposed by Kuorikoski et al. (2010), which requires that if a result R is derived with the core C and an auxiliary A_i, then learning this should not alter one's belief about whether R can be derived with C and another auxiliary A_j. Harris argues that this condition is violated because the core (e.g., C in Eq. 2) appears in both derivations, thereby increasing modellers' expectation that it is essential for deriving R. Her argument is cogent, but only if the DCR is accepted.

2.2 The tacking threat and genuine confirmation

While DDR can also be justified on intuitive grounds, I now briefly justify both rules using conceptual considerations. A more extensive formal treatment, drawing on genuine confirmation, is provided in the online appendix at: https://www.mv.helsinki.fi/home/alehtine/publications/EJPSappendix.pdf.

It is easy to construct an apparent counterexample to the rationality of DCR. Given a model

$$M_1 = (CA_1A_2A_3) \vdash R, \tag{4}$$

we can tack on any irrelevant assumption A (e.g. "the moon is made of green cheese") to obtain

$$M_1' = (CA_1A_2A_3A) \vdash R. \tag{5}$$

By monotonicity of entailment, DCR would treat this as confirming, even though A is known to be irrelevant. This is simply the tacking—or conjunction—problem applied to DCR, and a reviewer rightly raised it. However, the problem extends to indirect confirmation. A piece of evidence E confirms a component X indirectly when X does not entail E, yet E still supports X. Since the next section analyses such confirmation relations, it is necessary to address the potential problem. The solution is to use DDR to assign $p(A \vdash_C R|B)=0$: DDR blocks tacking by recognising that irrelevant assumptions are not necessary for deriving R. Schurz's (2014) condition makes this precise: confirmation spreads from a conjunction to one of its content elements only if that element is necessary for making the evidence likely. Since M_1 makes R at least as probable as $M_1{}'$, A fails this necessity test and is not confirmed.

Thus $p(X \vdash_C R|B)$ expresses a belief about whether X is necessary for deriving R, and only such elements can be genuinely confirmed. Schurz's condition thereby blocks tacking and secures the rationality of DCR, DDR, and indirect confirmation. His condition functions as a constraint on rational belief: it is not enough to propose a probability distribution that violates it—one must also show that such a distribution would be rational.

2.3 Preconditions for the use of the inferential rules and the strength of non-empirical confirmation

All accounts of robustness presuppose some degree of model variation and often a measure of independence among the varied elements. Although the inferential rules are not formulated as explicit independence conditions, together with their preconditions they embody a degree of diversity and weak independence similar to Schupbach's (2018) RA-diversity. This subsection articulates those preconditions, compares them with Schupbach's requirements, and clarifies where the inferential rules diverge from his account.

The key issue is that if robustness is demonstrated only with respect to auxiliaries that modellers already expect to be irrelevant, any resulting change in p(R|C) will be negligible. The preconditions are needed to ensure the relevance of the inferential rules.

Since some changes in assumptions do not affect conditional probabilities, additional preconditions for applying the inferential rules are needed. As noted earlier, these rules do not fully determine the relevant background knowledge about conditional dependencies among model elements and their results; the content of the assumptions and the nature of the results also matter. Modellers' background information may already indicate that an auxiliary A_i cannot influence the result R, even if A_i appears in every known derivation of the robust result. In such cases, the background knowledge supports the inference that the corresponding conditional probability either remains unchanged or alters insignificantly. For such an A_i, $p(A_i \vdash_C R|B_0) \approx 0$ beforehand and $p(A_i \vdash_C R|B_1)=0$ after deriving R without it. If robustness eliminates only such an assumption, the resulting increase in p(R|C) will be negligible.

Let U^{R_C} denote the set of elements known to contribute to deriving R within a model family associated with the robust theorem 'cp, if C then R'. Thus U^{R_C} comprises those elements which, according to the modeller's background knowledge, have been used in deriving R in a model containing C. When the robustness of R is established, some of these elements are shown to be superfluous.

Since the inferential rules operate on what modellers know and can identify, the elements of U^{R_C} are precisely those identified by the modeller. In (2), $U^{R_C}=\{C,A_1,A_2,A_3\}$, and in (3), $U^{R_C}=\{C,A_1,A_2,A_3,A_4,A_5\}$. A demonstration of robustness justifiably increases p(R|C,B) only if at least one auxiliary assumption in U^{R_C} that might have contributed to R is shown to be irrelevant

to the robust result.

Note that U^{R_C} does not include factors which, according to modellers' background knowledge, could in reality contribute to R but which were not included in the model family associated with the robust theorem concerning C and R. R could be causally overdetermined by X and other factors omitted from the model family. This is why rule DDR does not necessarily yield $p(X \vdash_C R|B_1)=0$, but instead merely decreases it.

For example, institutionalised racism in city planning provided an independent explanation for racial segregation when Schelling introduced his checkerboard model. If both mild same-race preferences and institutionalised racism are independently sufficient to produce segregation, then segregation is causally overdetermined by these mechanisms. Deriving the segregation result without invoking institutionalised racism reduces the probability of segregation given that factor, but does not eliminate its possible causal contribution.

If DDR is reformulated to apply only to the elements of U^{R_C}, it can be stated in a stronger form:

Rule DDR^C: If $X \in U^{R_C}$, and a result R is derived using a conjunction of identified elements that does not include X, then $p(X \vdash_C R|B_1)=0$.

We are now ready to state the two preconditions.

Precondition 1: No component X used to derive R may be known to entail R on its own: $p(R|X,B_0)<1$ for all $X \in U^{R_C}$.

Precondition 2: The Derivational Confirmation Rule operates only if there exists at least one auxiliary $A_i \in U^{R_C}$ such that $p(A_i \vdash_C R|B_0)>0$ and $p(A_i \vdash_C R|B_1)=0$.

The first precondition is not especially restrictive, as it is almost always met in real cases of robustness. The second precondition is more substantive: it requires that at least some auxiliaries shown by robustness to be irrelevant were not already known to be entirely irrelevant beforehand (i.e., $p(A_i \vdash_C R)=0$ prior to the derivation). Saying that X is not known to entail R on its own means that R is not known to be a deductive consequence of X. Modellers may know, for example, that $(XA_3A_5) \vdash R$ and $(XA_3) \vdash R$, but not that $X \vdash R$. Put differently, X would be known to entail R by itself only if, for all conceivable auxiliaries: $p(A_i \vdash_C R) \approx 0$.

The reason the preconditions for the inferential rules must differ from Schupbach's account is somewhat intricate, so I will first outline the core problem. Schupbach assumes that if a result R is explainable by a target hypothesis H, it will be detected with near-unit probability after several prior detections of R. Yet such a high probability could arise only if the new derivation of R eliminates assumptions that modellers already expected to be irrelevant (i.e., for which $p(A_i \vdash_C R) \approx 0$). Schupbach quantifies explanatoriness using a probabilistic notion that renders explanatoriness and expectedness indistinguishable. Since this measure is assessed prior to the derivation, it implies that the expected probability of detecting the robust result is already close to one before the detection occurs. Such an expectation can approach unity only if the components varied across derivations are already believed to differ minimally—or if all auxiliaries that change between derivations are assumed irrelevant from the outset.

If a modeller is genuinely uncertain whether a false auxiliary assumption is necessary for obtaining a result, this very uncertainty implies, ipso facto, that she does *not* expect the robust result to be derivable with near-unit probability after changing that assumption while keeping the core fixed. Robustness gains epistemic relevance only when it shows that at least some assumptions about which modellers hold such uncertainty are, in fact, irrelevant. Accordingly, the modeller's prior expectation that the robust result could be derived under these changes cannot itself be close to unity.

The inferential rules avoid this problem in a straightforward way. Their preconditions incorporate expectations about derivability, but they do not require that the rules apply only when all auxiliaries shown to be irrelevant by robustness were already expected to be irrelevant.

Let us now examine Schupbach's account more closely. It imposes a requirement similar to Precondition 2, stipulating that an alternative hypothesis must be capable of explaining previous detections of the robust result. Robustness Analysis (RA) diversity is defined as follows:

> Means of detecting R are RA-diverse with respect to potential explanation (target hypothesis) H and its competitors to the extent that their detections (R_1,R_2,...,R_n) can be put into a sequence for which any member is explanatorily discriminating between H and some competing explanation(s) not yet ruled out by the prior members of that sequence (Schupbach 2018, p. 288).

In Schupbach's framework, altering an auxiliary assumption can yield the

requisite weak independence, provided that the change enables explanatory discrimination between competing hypotheses. The inferential rules proposed here recognise a similar role for such auxiliary variation but are formulated in terms of expectedness rather than explanatoriness, since Schupbach's specific formulation of explanatory discrimination is rejected. The basic intuition behind explanatory discrimination is sound; the problem lies in Schupbach's excessively strong assumption needed to prove that robustness confirms. Like RA-diversity, DCR requires variation among the elements used to derive the robust result and interprets the epistemic gain from robustness as the elimination of false auxiliaries. The point is not to deny the relevance of explanatory considerations, but to note that, for any formal proof of confirmatory robustness, such considerations must be mediated through the expectedness of results (see also Roche & Sober 2013).

Since Schupbach uses the Volterra principle to characterise the hypothesis H in the modelling context, the most natural interpretation of H's competitor H´ is that some auxiliary A_i causes R. Schupbach denotes the conjunction of prior *detections* of the result $R_1\&R_2\&...\&R_{n-1}$ by **E**, and requires that both H and H' explain **E**. In modelling, 'detection' corresponds to deriving the result. More formally, the *success condition* states that $\varepsilon(\mathbf{E},H)=\varepsilon(\mathbf{E},H')>0$, where ε is the probabilistic measure of explanatory power defined by Schupbach and Sprenger (2011):

$$\varepsilon(e,h)=\frac{p(h|e)-p(h|\sim e)}{p(h|e)+p(h|\sim e)},$$

where h is explanans and e explanandum. The success condition parallels precondition 2.

Schupbach defines explanatory discrimination as the requirement that there exists an explanatorily discriminating means of potentially detecting R_n. In other words, H should strongly explain detecting the result once again (R_n), while H´ should strongly explain not detecting it ($\sim R_n$). Thus, the discrimination condition demands that H explains the renewed detection of the robust result, whereas H´ does not. Formally, this is expressed using the probabilistic measure of explanatory power ε: $\varepsilon(R_n,H|\mathbf{E})\approx 1$, and $\varepsilon(\sim R_n,H'|\mathbf{E})\approx 1$.[3]In the modelling context, RA-diversity can be interpreted

[3]Substituting H for h and R_n for e in ε(e,h), and conditioning on E, yields $\varepsilon(R_n,H|\mathbf{E})$. Schupbach's (2018) framework does not explicitly include an assumption corresponding to rule DDR^C. However, eliminative reasoning requires such an assumption, and Schupbach's stipulation that a hypothesis is discarded if it fails to explain the n nth detection of R

as follows: a series of models, each entailing the robust result R, are RA-diverse if every derivation shows that some auxiliary used in an earlier derivation is unnecessary, since R can be derived without it.

Let us now see why the inferential rules do not rely on explanatory discrimination. The concepts ε(e,h) and the conditional probability p(e|h) are closely interrelated (Schupbach 2017), such that $\varepsilon(\mathrm{R}_n,\mathrm{H}|\mathbf{E})$ cannot increase unless $\mathrm{p}(\mathrm{R}_n|\mathrm{H},\mathbf{E})$ also increases, and vice versa. Hence, when ε is used, explanatoriness and expectedness become indistinguishable. Unlike in experimental contexts, where hypotheses typically concern causal relations, in modelling both H and H´ address whether a component (C or A_i) is necessary for deriving the robust result. If an auxiliary A_i is shown to be irrelevant for R by deriving R from a model lacking A_i, then the competing hypothesis H´—claiming that A_i causes R—cannot explain why R_n can still be derived. This explains Schupbach's assumption that $\varepsilon(\sim\mathrm{R}_n, \mathrm{H}' \mid \mathrm{E}) \approx 1$.

A derivational history of the robust result R in a sequence of n−1 models, **E**, that satisfies the success condition could look as follows:

$$\begin{array}{cc} R_1: & M_1 = (CA_1A_2A_3A_4)\vdash R \\ R_2: & M_2 = (CA_1A_2A_3A_5)\vdash R \\ \vdots & \vdots \\ R_{\text{n-1}}: & M_{\text{n-1}} = (CA_1A_2A_3A_6A_7)\vdash R \end{array} \tag{6}$$

The explanatory power $\varepsilon(\mathrm{R}_n,\mathrm{H}|\mathbf{E})$ is then evaluated *prior to* conducting the next derivation:

$$R_n : M_n = (CA_1'A_2A_3A_4)\vdash R. \tag{7}$$

Schupbach assumes that when RA-diversity is satisfied, using C to derive the robust result R increases $\varepsilon(\mathrm{R}_n,\mathrm{H}|\mathbf{E})$. He further assumes that ε not only increases but approaches unity—$\varepsilon(\mathrm{R}_n,\mathrm{H}|\mathbf{E})\approx 1$—when an auxiliary A_1 initially used to derive R in M_1,...,M_{n-1} is later replaced byA_1'. The inference that $\mathrm{p}(\mathrm{R}_n|\mathrm{H},\mathbf{E})\approx 1$ implies that detecting R_n is virtually certain, given the robust theorem and prior derivations of R(**E**). This assumption is overly strong, as it limits the account's applicability to trivial cases.

To see why, consider the following question: if $\varepsilon(\mathrm{R}_n,\mathrm{H}|\mathbf{E})\approx 1$ results from C participating in the derivation of R in M_n while A_1 does not, did the preceding derivations—$\mathrm{R}_1,\mathrm{R}_2,...,\mathrm{R}_{n-1}$—already drive the explanatory powers $\varepsilon(\mathrm{R}_{n-1},\mathrm{H}|\mathbf{E})$, $\varepsilon(\mathrm{R}_{n-2},\mathrm{H}|\mathbf{E})$, and so on, close to unity? In a framework

serves the same function. This is why he assumes that $\varepsilon(\mathrm{R}_n,\mathrm{H'}|\mathbf{E})=0$.

where explanatoriness and expectedness coincide, this is equivalent to asking whether the earlier detections themselves were expected with near-unit probability—that is, whether p(R_{n-1}|H,**E**), p(R_{n-2}|H,**E**),... were already close to one.

McLoone et al. (2025) note that Schupbach's account violates logical omniscience (LO), and that under LO it would be impossible to fail to detect a deductive consequence of any conjunction of elements from which R is derivable. This, in turn, implies that ε could never change, since the conditional probabilities $p(R|H, \boldsymbol{E})$ and the detection probabilities p(R_{n-1}|H,**E**), p(R_{n-2}|H,**E**), and so forth would already be unitary. Under LO, therefore, there would be no need for a sequence of derivations—or for studying robustness at all—because every value of ε would simply equal one. This is correct, but it only shows that, under LO, no learning from robustness is possible: modellers would already know all derivational results in advance.

Schupbach need not assume logical omniscience, in which case the values of ε—and the associated expectations—are specified prior to the actual derivation. By contrast, the inferential rules change the conditional probability p(R|C) only after a derivation using C and a new set of auxiliaries has been performed. This distinction is crucial, since a change in $\varepsilon(R_n,H|\mathbf{E})$ is indistinguishable from a change in expectedness, $p(R_n|H, \boldsymbol{E})$.

To explain how it is possible that the previous detections were also expected with near-unit probability, consider the same question without assuming logical omniscience. Recall that Schupbach distinguishes the robust result R from its individual detections $R_1,...,R_n$. Since the hypothesis H concerns the core C and its relation to R, we can express p(H) = p(R|C). The inferential rules apply directly to R, C, and the auxiliaries A_i, whereas Schupbach's framework concerns the probability of detecting the result—R_n—given earlier detections **E**=$R_1,...,R_{n-1}$ and H. Thus, even if $\varepsilon(R_n,H|\mathbf{E})$, $\varepsilon(R_{n-1},H|\mathbf{E})$, and so on are close to unity, p(R|C) need not be. Modellers may still regard the robust theorem as open to (non-empirical) confirmation, even when the probability of detecting the result in the next model is already high.

This raises the question: under what circumstances could this occur? Even if $p(R_{n-1}|H, \boldsymbol{E})$, $p(R_{n-2}|H, \boldsymbol{E})$, and so on are close to unity, p(R|C) may still be considerably lower if some auxiliaries that might be necessary for deriving R have not yet been tested for robustness. This situation is perfectly coherent. However, it poses a surprising problem for Schupbach's assumption that $\varepsilon(R_n,H|\mathbf{E})\approx 1$.

To see the problem, consider what could justify believing that the robust

result will be derivable from a new set of assumptions with probability close to one, *given* **E**. This could only be because the previous derivations **E** relied on assumptions almost identical to those in the new model. In particular, any difference between two derivations would have to involve auxiliaries already expected to be nearly irrelevant to deriving R. If this were not the case—if changing an auxiliary (such as A_3 in our example) were thought likely to affect derivability—then the modeller could not simultaneously expect $p(R_n|H, \boldsymbol{E}) \approx 1$ for a derivation that alters that assumption. In short, near-unit expectations are possible only when auxiliary variation is already believed to be irrelevant, which makes the corresponding robustness result epistemically trivial.

Because explanatory discrimination cannot address auxiliaries that could be necessary for the robust result, any learning from robustness becomes trivial *at each stage*, leaving p(R|C) virtually unchanged. Conversely, if any of the earlier values—$\varepsilon(R_{n-1},H|\mathbf{E})$, $\varepsilon(R_{n-2},H|\mathbf{E})$, and so on—were not close to unity, Schupbach's proof of confirmatory robustness (i.e., that $p(H|R_{n-1}, \boldsymbol{E}) > p(H|R_{n-2}, \boldsymbol{E})$) could not hold for that stage. In short, insofar as Schupbach's account assumes $\varepsilon(R_n,H|\mathbf{E})\approx 1$, it applies only to cases where robustness rules out auxiliaries that were already known to be trivial.

Admittedly, the expectation $p(R_{n-1}|H, \boldsymbol{E})$ could approach unity if p(R|C) were already close to one. Yet the nearer it comes to unity, the closer one is to violating Precondition 1, which is meant to exclude this second form of irrelevance. Some earlier accounts (e.g., Weisberg 2006; Kuorikoski et al. 2010, Dethier 2024) explicitly adopt such an assumption. It is, however, unwise to rely on it when arguing for the confirmatory value of robustness, since increases in p(R|C) do not require that this probability be high to begin with. The inferential rules certainly do not depend on this assumption.

It is conceivable that Schupbach's proof could be reformulated under the weaker assumption that $\varepsilon(R_n, H|\boldsymbol{E}) > \varepsilon(R_{n-1}, H|\boldsymbol{E})$. Yet philosophers' beliefs about whether such a proof can be given also violate logical omniscience: we cannot know whether such a proof exists until someone actually constructs it—or demonstrates, by impossibility, that none can be.

It is instructive to recall a familiar criticism of robustness: that it yields no epistemic gain unless all possible idealisations are neutralised (e.g., Odenbaugh & Alexandrova 2011; Harris 2021). Harris further contends that robustness does not necessarily warrant transferring claims from the model world to the real world. Such concerns are understandable if Schupbach's assumption that $p(R_n|H, \boldsymbol{E}) \approx 1$ is compatible with p(R|C) being substan-

tially lower (even though Harris does not explicitly target Schupbach). In particular, the epistemic payoff is negligible if robustness merely shows the irrelevance of assumptions already expected to be irrelevant.

The preconditions for the inferential rules also reveal what the epistemic relevance of robustness depends on: the difference between the expected derivability of a result and the established knowledge of derivability once robustness is demonstrated. Crucially, modellers need neither to know all possible auxiliaries that could be varied nor to verify that the true auxiliary is among those considered for robustness to yield incremental epistemic gain. Deriving the robust result once more, as in (7), provides non-empirical confirmation even if some false auxiliaries remain. The relevant assumptions are those in U^{R_C}, not the set of all conceivable ones.

When auxiliaries are nearly true, modellers may feel less concerned about their potential to generate erroneous results. Yet, in assessing how p(R|C,B), or p(R|A_i,B) changes with the robustness of R, modellers focus on the derivational relationship between C or $\mathrm{A_i}$ and R, not on their truth values. The increase in non-empirical confirmation of the robust theorem depends on the irrelevance of certain auxiliaries, not on their truth. By contrast, the truth values of auxiliaries matter when evaluating the absolute degree of confirmation of the robust result, especially when it lacks direct empirical support. In short, establishing p(R|C,$\mathrm{B_1}$)>p(R|C,$\mathrm{B_0}$) requires believing that the core C has become more relevant following the demonstration of robustness, and that at least some auxiliaries are irrelevant to R.

While preconditions 1 and 2 specify when the inferential rules can be applied, reformulating them also indicates how much non-empirical confirmation these rules can yield. The underlying intuition is that the more strongly a modeller initially believes that certain false auxiliaries might be necessary for the robust result R, and the less confidence they have that C is necessary, the greater the increase in p(R|C) once R is shown to be robust.

For instance, if a modeller begins with $\mathrm{M_1}$=(C$\mathrm{A_1A_2A_3}$)⊢R, believing that auxiliary A_1 is essential, then discovering through $\mathrm{M_2}$=(C$\mathrm{A_2A_4A_5}$)⊢R that A_1 is unnecessary increases confidence that the remaining elements are genuinely needed. Conversely, if a robust theorem is already well established, additional demonstrations of its robustness can contribute only marginally to its confirmation. The strength rules are as follows.

Strength rule 1: The smaller p(R|X,$\mathrm{B_0}$) is initially, the more demonstrating the robustness of R with a set containing X can increase it.

Strength rule 2: The greater the number of assumptions $A_i \in U^{R_C}$ for which $p(A_i \vdash_C R|B_0) > 0$, and the greater the inequality $p(A_i \vdash_C R|B_0) > p(A_i \vdash_C R|B_1)$ for auxiliaries shown to be irrelevant, the larger the potential increase in $p(R|X,B)$.

In exceptional cases—such as Einstein's derivation of Mercury's perihelion precession from general relativity—a single derivation can dramatically raise the relevant conditional probability. In more typical instances of derivational robustness, however, one, two, or even several derivations do not raise $p(R|C, B_1)$ to unity. Further demonstrations of robustness continue to increase this probability, but each successive derivation yields a progressively smaller quantitative gain.

The key difference from Schupbach's account is that the inferential rules do not assume $\varepsilon(R_n,H|\mathbf{E}) \approx 1$, since this assumption effectively requires that the inequalities $p(A_i \vdash_C R) > 0$ and $p(R|A_i,B_1) < p(R|A_i,B_0)$ be negligible. Schupbach's framework applies only when, for for $A_i \in U^{R_C}$: $p(A_i \vdash_C R) \gtrapprox 0$ and $p(R|A_i,B_1) \lessapprox p(R|A_i,B_0)$. By contrast, the inferential rules also apply when $p(A_i \vdash_C R) \neq 0$ for the auxiliaries being modified, and the strength rules specify how the quantitative increase in $p(R|X)$ depends on these probabilities.

3 Inferential rules for experimental robustness

The inferential rules discussed above concern derivational robustness, but analogous—though slightly different—rules can be formulated for experiments and measurements. To clarify these distinctions, let us examine the main differences between experiments and models.

First, unlike modellers, experimenters do not necessarily know all the components involved in their experiments. In this context, the elements in parentheses in equations (2) and (3) represent the set of elements the experimenter recognises, rather than an exhaustive list of all operative components. Second, experimenters may fail to identify the core elements that generate their results—for example, Brown did not initially understand the cause of pollen movement in Brownian motion. Moreover, there may be no common elements shared across all experiments that yield the same result R. Third, while a degree of variety is typically required in both contexts, a substantial body of work (e.g., Bovens & Hartmann 2003; Claveau 2013; Osimani

& Landes 2023; see also Landes 2020) suggests that the variety-of-evidence thesis may fail if repeated experiments provide sufficiently strong information about experimental reliability. Finally, because R itself constitutes empirical evidence, the experimenter's conditional probabilities can be interpreted in terms of expected experimental results given those already observed. Relatedly, experimenters need not be assumed to violate logical omniscience. Consequently, since R is empirical, the conditional probability $p(R|B_1)$ is well defined.

Casini and Landes (2024) argue that a model's robustness can yield confirmation even without variation among auxiliaries, suggesting that results from the variety-of-evidence literature in experimentation apply, mutatis mutandis, to modelling. I disagree. Modelling differs from experimentation in a crucial respect: deriving R with M_1 after already establishing it in (2) provides no new information. An increase in $p(R|C,B_1)$ requires that the new set of auxiliaries contain at least one element not previously shown to be irrelevant. In experimentation, replication can be epistemically valuable; in modelling, by contrast, it is not—since deriving the same result from the same assumptions repeatedly adds no epistemically relevant information.[4]

In summary, once a result has been derived from a given set of elements, repeating the same derivation cannot enhance the reliability of the inference and is therefore epistemically uninformative. Although the inferential rules are similar across modelling, experimentation, and measurement, they are not identical. The variety-of-evidence thesis cannot fail in modelling in the same way it can in experimental or measurement contexts. Hence, Rule DCR includes the variety condition: "if this same set of elements has not already been used to derive R." When a result is derived for the first time, this condition is redundant but automatically satisfied. Accordingly, the corresponding experimental rules omit the variety requirement.

The rules for experimental robustness can thus be formulated as follows:

Experimental confirmation rule (ECR): If a result R is obtained in an experiment that includes component X, then the conditional probability $p(R|X)$ justifiably increases: $p(R|X,B_1)>p(R|X,B_0)$.

Experimental disconfirmation rule (EDR): If a result R is obtained in an experiment that excludes component X, then the conditional probabil-

[4]Computer simulations differ from analytical models in this respect, since running the same program on different physical machines may yield different results.

ity $p(R|X,B_1)$ decreases, or is set equal to $p(R|B_1)$.

The inferential rules apply, with slight modifications, to both experimental and derivational robustness, though their confirmatory roles differ. In experiments, the robust result itself constitutes empirical evidence, and robustness strengthens confirmation by increasing evidential variety. In modelling, by contrast, the result R may be theoretical, and robustness then provides only non-empirical confirmation. If R, or some related result, is supported by empirical evidence E such that $p(E|R)>p(E)$, derivational robustness can also yield empirical confirmation. An increase in $p(R|C)$ alone does not constitute empirical confirmation, though it is a necessary precondition for it in modelling. Empirical confirmation involves additional, indirect relations that seldom arise in experimentation or measurement, where robust results already serve as evidence.

Because an increase in $p(R|C)$ within modelling contexts does not yet amount to empirical confirmation, I develop an account that does—by reformulating and extending Lehtinen's account of indirect confirmation through robustness using the inferential rules.

4 Empirical confirmation from derivational robustness

In this section, I will provide a systematic account of empirical confirmation from derivational robustness. I will use Lloyd's (2015) account of 'model robustness' in climate models (see also O'Loughlin 2021) as a starting point for identifying the possible confirmation and robustness relations in a family of models. Given that her account is designed to apply to climate modelling, I will adopt her interpretations of the model components. Models dealing with complex systems often involve a multitude of variables and various types of relevant empirical evidence. Lloyd (2010) and Lehtinen (2018) discuss the following robust theorem: 'cp, if there is an increase in CO_2 forcing (C), the global mean surface temperature rises (R)' (see also Winsberg 2021).

Climate studies aim to determine future temperatures and understand how they depend on CO_2 forcing, which makes the robustness of this relationship crucial. A single model could be described as follows:

$$\begin{array}{lcccc} M_i = (C, A_i ...) \vdash & R_1 & R_2 & R_3 & ... \\ \top & | & | & | & \\ R_M, R_{M2}, R_{M3} & E & E_2 & E_3 & \end{array} \tag{8}$$

Here, C can be interpreted as CO_2 forcing, and A_1 as auxiliaries concerning, for example, cloud formation or different ice albedo. E represents data on global mean surface temperature, R the model's prediction of said temperature, and R_M the model's prediction of *future* global mean surface temperature. The difference between results like R_{M1}, R_{M2} and R_{M3} and results R_1, R_2 etc., is that the former lack direct supporting evidence, whereas the latter possess it. The vertical line represents the evidential relations. For example the line between R_1 and E means $p(R_1|E)>p(R)$.

Graph (9) provides a different representation of (8) in that results R_{M2}, R_{M3}, and R_2, R_3,... are omitted, while some of the model elements are displayed.

$$\begin{array}{lccccccc} M_i = & (C & A_1 & A_2 & A_3) & \vdash & R \\ \top & | & | & | & | & & | \\ R_M & E_C & E_{A1} & E_{A2} & E_{A3} & & E \end{array} \tag{9}$$

Let's assume that other models M_j in the same model family have comparable support for their auxiliaries:

$$\begin{array}{lccccccc} M_j = & (C & A_2 & A_4 & A_5) & \vdash & R \\ \top & | & | & | & | & & | \\ R_M & E_C & E_{A2} & E_{A4} & E_{A5} & & E \end{array} \tag{10}$$

Lloyd (2015) claims that model robustness confirms due to variety of direct and indirect evidence. An individual model M_i in a model family may possess direct empirical evidence for:

i) the core structure of the model C (E_C),

ii) the auxiliaries A_i (E_{Ai}),

iii) the result R (E),

iv) the relationship between C and R.

As O'Loughlin (2021) and Gluck (2023) note, Lloyd argues that all these pieces of evidence contribute to the confirmation of climate models as a whole, rather than providing a confirmation-theoretically strict notion of confirmation. While I appreciate Lloyd's effort to incorporate various kinds of empirical evidence into the discussion on robustness, her case-study-focused

approach does not explain in detail how robustness affects the different indirect confirmation relations into which such evidence can enter. For this reason, I will examine all the relationships in which these pieces of evidence may play a role. My aim is to determine whether these relations are relevant to robustness and, where robustness is pertinent, to clarify the nature of that relevance.

Lloyd (2015) mentions indirect but not direct evidence for C. In the context of climate modelling, such evidence consists of measured CO_2 emissions from the past. Given that there is little uncertainty about the accuracy of these measurements, and they are embedded as part of the climate models, it is natural for her to omit such evidence. I have included it here for the sake of completeness, and because in some other contexts the core is not modelled by parameterising it with empirical data.

Lloyd claims that empirical evidence E, E_2, E_3,...and all the E_{Ai} provide indirect evidence for the core structure C by increasing the variety of evidence. There are several ways to interpret what 'variety of evidence' means. Here, Lloyd appeals to a concept closest to Whewellian consilience: the variety of evidence stems from the differences in content and origin of E, E_2, E_3,..., E_C, E_{A1},... etc. Although Lloyd argues that variety of evidence is conceptually distinct from robustness, her notion of model robustness does not specify the exact relationship between the two. Since this variety of evidence would accrue to the core C even if there were only a single model M_1, and thus no robust results, the role of robustness in confirmation remains unclear in her account. On the other hand, the degree to which evidence E, E_2, E_3,... confirm C as well the future prediction about temperature R_M depends on the robustness of the results. However, such claims cannot be justified without resorting to an explicit account of indirect confirmation (Lehtinen 2016, 2018). He argued that derivational robustness confirms by strengthening the derivational links in indirect confirmation. In this paper, this somewhat ambiguous expression is replaced by an application of the inferential rules, providing a clearer account of the difference between non-empirical and empirical confirmation. Indirect confirmation arises from the fact that a result demonstrably depends on the same model components as another empirically confirmed model result.

Clearly, model robustness must be confirmatory if all the elements listed above and their possible relationships belong to model robustness. For instance, E_C undeniably confirms C, and E_{A1} confirms A_1 etc. I am also confident that climate researchers consider all of these confirmation relation-

ships to be relevant. Nevertheless, those who dispute that robustness offers confirmation might justifiably argue that Lloyd claims 'robustness confirms' merely because robustness has been redefined as a practice encompassing elements that are not related to robustness in terms of the sameness of results. A strict interpretation of robustness poses a more precise question: What is the contribution of the fact that R is derived from both M_i and M_j to the confirmation of something?

Winsberg (2021) attempts to articulate the distinction between Lloyd (2010, 2015) and Parker (2011) by claiming that Parker asks the precise question, whereas Lloyd does not (see also Gluck 2023). I will now provide answers to this precise question by applying the inferential rules to the indirect confirmation relations.

Since robustness of results R or R_M could not affect the conditional probabilities p(E |R), p(E_C|C), p(E_{A1}|A_1), etc., one can focus on the conditional probabilities between C and R or R_M, A_1 and R or R_M, and so on. Let 'X *R-confirms* Y' denote 'X confirms Y more than it would if some result were not (known to be) robust'.[5] I will make the following eight claims.

Evidence for the core structure of the model C (E_C) R-confirms

a1) R_M, iff the robustness of R_M justifies the inference p(R_M|C,B_1)>p(R_M|C,B_0), and it R-confirms

a2) R, iff the robustness of R justifies the inference p(R|C,B_1)>p(R|C,B_0),

Evidence for the auxiliary A_i (E_{Ai}) R-confirms

b1) R, iff the robustness or R justifies the inference p(R|A_i,B_1)>p(R|A_i,B_0). However, if a given auxiliary A_i does not take part in every derivation

[5]Given that there are several standard measures of the degree of confirmation—namely, the difference measure $D_D(H,E) = p(H|E)-p(H)$, the ratio measure $D_R(H,E)=p(H|E)/p(H)$, and the likelihood measure $D_L(H,E)=p(E|H)/p(E|\tilde{}H)$) — one might wonder whether R-confirmation behaves differently depending on the chosen measure. However, it is not necessary to distinguish between these measures here, because all three satisfy the following adequacy condition under standard assumptions: If two pieces of evidence E_1 and E_2 both confirm a hypothesis H (i.e., $p(H|E_1)>p(H)$ and $p(H|E_2)>p(H)$), and if $p(H|E_1)>p(H|E_2)$, then the degree of confirmation of H by E_1 exceeds that by E_2. This is because each of these measures is monotonic in the posterior probability p(H|E) when appropriate background parameters (e.g., priors or likelihoods) are held fixed. See Fitelson (1999) and Sprenger & Hartmann (2019, ch. 5–6) for detailed discussion.

of a robust result R, then $p(A_i \vdash_C R|B_1)=0$, and evidence for such an auxiliary cannot confirm the robust result.

b2) R_M, iff R_M's robustness justifies the inference $p(R_M|A_i,B_1)>p(R_M|A_i,B_0)$, and the robustness of R justifies the inference $p(R|A_i,B_1)>p(R|A_i,B_0)$. As in b1), if an auxiliary is replaced by another auxiliary in either derivation, evidence for it cannot confirm due to robustness.

Evidence for the result R (E) R-confirms

c1) the core C, or the hypothesis that C causes R, iff the robustness of R justifies the inference $p(C|R,B_1)>p(C|R,B_0)$.

c2) R_M iff the robustness of result R_M justifies the inference $(R_M|C,B_1)>p(R_M|C,B_0)$ and the robustness of result R justifies the inference $(R|C,B_1)>p(R|C,B_0)$.

d) Empirical evidence for the relationship between C and R confirms R, but the robustness of R could not affect its strength.

e) R-confirmation due to derivational robustness of results is non-monotonic.

A few comments about these claims. First, given that each confirmation relation is based on whether the robustness of a result changes the relevant conditional probabilities, there cannot be empirical R-confirmation without non-empirical confirmation from robustness. Second, the expression 'justifies the inference' above is short for 'applying the inferential rules justifies the inference'. It follows that none of these results are acceptable if my arguments for the inferential rules are not deemed acceptable. Third, all of these cases necessitate the ability to pinpoint which individual components within the models are being confirmed. Finally, all the claims pertain to indirect confirmation.

Lehtinen (2016, 2018) has already made similar arguments for b1, c1, and c2. Here, I aim to further clarify b1 in response to objections raised against Lehtinen's claim, and provide a more accessible account of c1 and c2. Lloyd (2015) discusses the overall confirmation of climate models and claims that it also contributes to the trustworthiness of future projections. I am proposing c2 as a possible interpretation of Lloyd's claim, and i consider it as the single most important form of empirical confirmation derived from robustness. In the context of climate science, for example, this concerns whether the robustness of current climate model ensembles indirectly confirms future temperature projections.

This paper precisely articulates the intuition underpinning indirect confirmation via robustness: if robustness shows that a result (R_M) depends on the same assumptions (the core C) as another empirically confirmed result (R), then R_M may be indirectly confirmed through empirical evidence for R. Robustness can reveal that the result R_M relies on assumptions not only better supported by evidence for R, but also more aligned with R than was indicated by the modellers' background knowledge (B_0), thus indirectly confirming it. I will demonstrate how this intuition can be precisely analysed by applying the inferential rules to the indirect confirmation relations relevant for derivational robustness.

Recall that, following Orzack and Sober's (1993) argument, derivational robustness can empirically confirm only indirectly. Although confirmational support is not generally transitive, indirect confirmation specifically requires the transmission of support along a chain. Before proceeding to the case-by-case analysis, it is important to emphasise a result from Shogenji (2017): while the transitivity of overall confirmation cannot be guaranteed, the indirect component of confirmation is transitive. That is, although other probabilistic dependencies may block overall confirmation from evidence E (here E, E_C, or E_{Ai}, depending on the case) to a target hypothesis H (here C, R_M or R), the support that E transmits via an intermediary hypothesis H´ (here C or R) is preserved. Since our analysis modifies only one conditional probability at a time, with all other background conditions held constant, the possibility of non-transitivity in overall confirmation does not undermine the validity of the results presented here.

4.1 a1) Evidence for the core (E_C) R-confirms R_M

To show that E_C confirms the result R_M, it is helpful to express models (9) and (10) as follows:

$$M_i = \begin{array}{cccc} (\mathrm{C} & A_1 & A_2 & A_3) \\ | & | & | & | \\ E_C & E_{A1} & E_{A2} & E_{A3} \end{array} \vdash R_M \qquad M_j = \begin{array}{cccc} (\mathrm{C} & A_2 & A_4 & A_5) \\ | & | & | & | \\ E_C & E_{A2} & E_{A4} & E_{A5} \end{array} \vdash R_M \tag{11}$$

Given that the two models contain different auxiliaries but share C, applying rules DCR and DDR to R_M and C justifies the inference $p(R_M|C,B_1)>p(R_M|C,B_0)$. This means that evidence E_C confirms R_M more strongly after demonstrating the robustness of R_M because the likelihood $p(E_C|C,)$ is unchanged when

the robustness of R is demonstrated in (10). Note that the robustness of R_M R-confirms itself. I reiterate, the robustness of result R_M empirically confirms R_M itself! Nonetheless, it lacks confirmation from any direct piece of evidence since it doesn't have any. Its R-confirmation is empirical, genuine, incremental, and indirect.

4.2 a2) Evidence for the core (E_C) R-confirms R

The argument for why the robustness of R confirms R itself is similar. Given that the two models contain different auxiliaries but share C, applying rules DCR and DDR justifies the inference that $p(R|C,B_1)>p(R|C,B_0)$.

4.3 b1), b2) Evidence for the auxiliaries (E_{Ai}) R-confirms R or R_M only if they are used in deriving the results in all models

Consider the case of direct empirical evidence for auxiliary A_1, after having derived R from model M_i but not from model M_j. Initially, it might seem plausible that A_1 is necessary for deriving R, so E_{Ai} indirectly supports R. However, the robustness of R—following its derivation from M_j, where A_1 is absent—demonstrates that A_1 is not necessary for deriving R. This removes any initial confirmation provided by E_{Ai}. Therefore, evidence for the auxiliaries A_i does not confirm the result R or R_M due to robustness unless the auxiliary is involved in the derivation of these results in *all* relevant models. When it becomes evident that an auxiliary like A_1 is irrelevant to the result R, empirical evidence for that auxiliary no longer contributes to the indirect confirmation of R. This is because the confirmation relation between E_{A1} and R is shown to be non-genuine. Suggesting that evidence for irrelevant auxiliaries confirms a robust result would imply that tacking irrelevant components to models is acceptable and that evidence for these components matters when evaluating results derived from these models. Therefore, applying rule DDR justifies the conclusion that R is not confirmed by E_{A1}.

This argument was originally presented by Lehtinen (2018), and I have rephrased it here in light of objections from O'Loughlin & Li (2022. see also Fuller & Schulz 2021). Lehtinen's initial statement might have appeared misleading because it seemed to suggest that the auxiliary assumptions had to be mutually incompatible. Indeed, the literature on climate modeling,

which was his case study, acknowledges the potential for such incompatibility. However, contrary to what O'Loughlin & Li (2022) claim, this argument does not depend on whether the auxiliaries have dichotomous truth values or are mutually incompatible. The key issue is whether these assumptions are relevant for deriving the result R.

Even if auxiliary assumptions share common content, that commonality is relevant to the robust result R. However, in such cases, the shared content must be explicitly represented when evaluating its relevance to robust results. I have depicted the auxiliaries as having distinct but not necessarily incompatible content, with the empirical evidence E_{A1} concerning that specific content. To the best of my knowledge, different cloud formation modules, for example, do have overlapping content. Therefore, if the empirical evidence concerns entire modules, it should not be represented like E_{Ai} in diagrams like (9) and (10). A more accurate representation would involve replacing A_1 and A_4 with modules O_1=($\mathrm{A}_1\mathrm{A}_x\mathrm{A}_y\mathrm{A}_z$,...) and O_4=($\mathrm{A}_4\mathrm{A}_x\mathrm{A}_y\mathrm{A}_z$,...), where common elements A_x, A_y, A_z,...are explicitly shown. If no specific evidence pertains to A_1 or A_4 another example should be used to illustrate the point.

Since auxiliary A_2 takes part in both derivations (9) and (10), the empirical evidence E_{A2} genuinely and indirectly confirms both R and R_M. This demonstrates the importance of distinguishing between absolute and incremental confirmation when discussing the confirmation of a robust result. If any auxiliary that is essential to all derivations of a robust result is disconfirmed by direct evidence, then that evidence also disconfirms the robust result. This gives weight to the criticisms of robustness in scientific modelling, which argue that robustness does not guarantee truth in many cases. The reason for this is that the remaining shared false auxiliaries could still be responsible for a robust result.

However, the incremental confirmation provided by robustness is still relevant. The robustness of R brings incremental confirmation from empirical evidence for an auxiliary A_i if A_i is necessary for deriving R. This holds true even though robustness may not guarantee a high absolute probability of R being true. Lloyd, O'Loughlin, and Li are correct to emphasize the significance of direct evidence for auxiliaries when assessing the absolute confirmation of a robust result. But the logic underlying genuine confirmation, as expressed by rule DDR, indicates that such evidence incrementally confirms the robust result R due to robustness only if the auxiliaries are shared among all the models that derive R.

4.4 c1) Evidence for the robust result (E) R-confirms the core C

The likelihood p(E|C) represents how closely the core C is related to the evidence E. In this setting, we can decompose p(E|C) into p(E|R), the probability of the evidence given the result R, and p(R|C), the probability of the result given the core structure. Since p(E|R) is determined by various data-to-phenomena inferences (Bogen and Woodward 1988) and is unaffected by the robustness of R, the conditional probabilities p(E|R) and p(R|C) are independent of each other. Consider the consequences of demonstrating the robustness of R by learning (10) when (9) is already at hand. Since C takes part in both derivations, applying rule DCR imples that p(R|C) increases. Here robustness strengthens the connection between R and C. As a result, p(E|C) increases, even though p(E|R) is unchanged. Thus, by increasing the probability of R given C, robustness indirectly increases the probability of the old evidence E given C. This reasoning explains how evidence for the robust result R can R-confirm the core C, despite E being old evidence. Robustness strengthens the logical connection between the core and the result, leading to an increase in the confirmation of C from the evidence E.

According to Bayes' theorem, the posterior probability of C given E and B_1 is $p(C|E,B_1)= \frac{p(E|C,B_1)p(C,B_1)}{p(E,B_1)}$. Since E is old evidence, we assume $p(E,B_1)=p(E,B_0)=1$. Additionally, it is reasonable to assume that $p(C|B_1)\geq p(C|B_0)$. From the fact that $p(E|C,B_1)\geq p(E|C,B_0)$ it follows that $p(C|E,B_1)>p(C|E,B_0)$. Therefore, since demonstrating robustness shows that C is more closely connected to R than previously thought, evidence E for R indirectly R-confirms C.

Casini and Landes (2024) discuss a case in which p(E|R) and p(R|C) are not distinguished from each other. Instead, they assume that p(E|C) increases due to robustness. By Bayes' theorem, C is confirmed by E in this simplified case too.

4.5 c2) Evidence for the robust result (E) R-confirms result $\mathbf{R}_M$

Given that p(E|R) is fixed, the goal here is to show that $p(R_M|R,B_1)>p(R_M|R,B_0)$. The intuition is that R_M is genuinely indirectly confirmed by E if it depends on the same model components as the confirmed result R. The robustness

of R R-confirms R_M because the application of the inferential rules demonstrates that these dependency relations are more secure than they would be without robustness. Unlike other confirmation relations, this one requires that both results R and R_M be robust. To see why, consider a counterfactual scenario in which one of the results is not robust. Suppose climate modellers run the first model only until the present time and do not draw any conclusions about the future. Further, assume that all the auxiliaries involved are disconfirmed by the direct evidence:

$$\begin{array}{lccccccc} M_i = & (\mathrm{C} & A_1 & A_2 & A_3) & \vdash & R \\ & | & | & | & | & & | \\ & E_C & \sim E_{A1} & \sim E_{A2} & \sim E_{A3} & & E \end{array} \tag{12}$$

Now consider that model M_j is used to derive results for both the present and the future:

$$\begin{array}{lccccccc} M_j = & (\mathrm{C} & A_2 & A_4 & A_5) & \vdash & R \\ \top & | & | & | & | & & | \\ R_M & E_C & \sim E_{A2} & \sim E_{A4} & \sim E_{A5} & & E \end{array} \tag{13}$$

Although applying DCR justifies the inference $\mathrm{p(R|C,B_1)>p(R|C,B_0)}$, there remains some ambiguity regarding whether E genuinely indirectly confirms R_M. This arises because it is unclear whether the components confirmed by applying rule DCR (namely C and A_2) are indeed required for deriving R_M from M_j. It could well be the case that R_M is actually dependent on auxiliaries A_4 or A_5 rather than C and A_2, meaning that the evidence E does not contribute to confirming R_M. Given that these auxiliaries are known to be false, R_M could be highly suspect. Moreover, if it were found that $(\mathrm{A_1A_3})\nvdash R_M$, then, applying rule DDR to this background information B, $\mathrm{p}(C\vdash_C \mathrm{R|B})=\mathrm{p}(A_2\vdash_C \mathrm{R|B})=0$ and hence $\mathrm{p}(R_M\vdash_C \mathrm{E|B_2})=0$ because A_1 and A_3 would be shown to be relevant to R_M, but irrelevant to the confirmed result R and, consequently, to the confirming evidence E. However, if instead R_M is now shown to be robust by deriving it from M_i, we get the following:

$$\begin{array}{lccccccc} M_i = & (\mathrm{C} & A_1 & A_2 & A_3) & \vdash & R \\ \top & | & | & | & | & & | \\ R_M & E_C & \sim E_{A1} & \sim E_{A2} & \sim E_{A3} & & E \end{array} \tag{14}$$

In this scenario, R_M becomes indirectly R-confirmed by E, as it is shown to depend, via an application of the inferential rules, on the components C

and A_2, which are genuinely confirmed by E. According to DCR, since C and A_2 participate in deriving R_M in (13), using these same components to derive R_M in another model that introduces some new auxiliaries and excludes others, as in (14), increases $p(R_M|CA_2)$. Consequently, given the background knowledge B_2 that arises from deriving (14), we have $p(R_M|CA_2,B_2)>p(R_M|CA_2,B_1)$. Since B_1 already implies that C&A_2 is R-confirmed by the robustness of R: $p(CA_2|E,B_1)>p(CA_2|E,B_0)$, combining these consequences of the robustness of R and of R_M yields $p(R_M|E,B_2)>p(R_M|E,B_0)$. Consequently, R_M is genuinely indirectly confirmed by its own robustness once the robustness of R is established. Once again, the robustness of a result confirms itself. This increased confirmation results from showing that the confirmation E confers on C&A_2 (via R) is more likely to be genuine, and that C&A_2 is more likely to genuinely confirm R_M. The application of rules DCR and DDR demonstrates how robustness strengthens the logical connections within the indirect confirmation structure. The graph (15) illustrates this indirect confirmation structure, with the symbols $\vdash_{\mathbf{c}}$ and $\top_{\mathbf{c}}$ indicating that the logical links have become stronger.

$$\begin{array}{ccc} (CA_2) & \vdash_{\mathbf{c}} & R \\ \top_{\mathbf{c}} & & \top \\ R_{M1} & & E \end{array} \tag{15}$$

Now, let's interpret this result in the context of climate models. The consistency between model-predicted and observed temperatures—expressed as $p(R)<p(R|E)$ —indirectly confirms the common prediction of the model family, namely the rise in GMST in the future (R_M). This indirect confirmation structure was already established in equation(13). However, robustness has strengthened these connections by strengthening the two robust theorems.

Consider that, in our example, the confirmation from evidence E is allocated to the conjunction of C and A_2 instead of C alone. If we fail to establish the irrelevance of every auxiliary, the absolute confirmation of the robust result or the robust theorem (if C, then R or R_M) may remain low, even after demonstrating the robustness of R or R_M. However, the fact that demonstrating robustness has not eliminated all the auxiliaries does not imply a complete lack of confirmation.

McLoone (2025) argues that components known to be false cannot be confirmed and, therefore, do not contribute to indirect confirmation. The falsity of A_2 leads to $p(CA_2)=0$. Additionally, according to Bayes' theorem, we have $p(CA_2|E,B_1)=\frac{p(E|CA_2,B_1)p(CA_2)}{p(E)}=0$. More generally, adopting this perspective

implies that models can never be confirmed, as they always involve idealisations that have a zero prior probability of being true (see Shaffer 2012, Sayan 2005). However, this viewpoint is unnecessarily restrictive, as it may be rational for modellers to take into account background information not obtained from formally demonstrating the irrelevance of auxiliaries, as we have seen. For instance, their background information might suggest that it is highly unlikely for A_2 to be necessary for deriving R or R_M, despite its consistent inclusion in these derivations due to its content. In such cases, modellers may consider the possibility of showing, for example, the following:

$$\begin{array}{ccccccc} M_k = & (\mathrm{C} & A_1 & A_3 & A_4) & \vdash & R \\ \top & | & | & | & | & & | \\ R_M & E_C & \sim E_{A1} & \sim E_{A3} & \sim E_{A4} & & E \end{array}, \tag{16}$$

even before the actual derivation is conducted. It would then be rational to infer that the robustness of the two results indirectly R-confirm R_M because $\mathrm{p}(\mathrm{R}_M|\mathrm{C},B_2)>\mathrm{p}(\mathrm{R}_M|\mathrm{C},B_1)$ even though a small probability remains that A_2 could be responsible for these results. For such inferences, the degree to which the auxiliary deviates from the truth is of little consequence; what truly matters is the possibility that it could be required in deriving the robust results. I therefore suggest that while the inability to eliminate all false auxiliaries diminishes the extent to which robustness indirectly confirms the robust result R_M (or a robust theorem), it does not negate the confirmation entirely. I acknowledge that I am only proposing a preliminary outline of how an account of confirming models with empirical evidence might take shape, and that a more comprehensive analysis is necessary. Nevertheless, even a sketch could prove more useful than outright denying that models can be confirmed due to the zero-probability problem.

4.6 Evidence for the connection between C and R

Direct empirical evidence for the connection between C and R can be obtained, for instance, if C is found to be correlated with R in a statistical test or through experiments, as noted by Lloyd, O'loughlin, and Li.[6] While such

[6]A terminological clarification is necessary here. O'loughlin and Li (2022) assert that the 'causal core' in climate models consists of the 'relationship between CO_2 and temperature'. In the context of this paper, this concept corresponds to the robust theorem 'ceteris paribus, if C then R', rather than a specific set of model components C. Thus, when they refer to 'evidence for the core,' they mean evidence for the connection between C and R,

evidence is undeniably important for evaluating the absolute confirmation of the robust theorem, the robustness of R does not influence it in any way. There can also be significant derivational evidence for this connection based on non-robustness. In climate science, modellers aim to test the robustness of the connection between the core (CO_2 forcing) and the robust result R (the observed global temperature rise) using 'control runs' that exclude the core from the climate model. These control runs typically demonstrate that the observed global temperature increase of about 1 °C cannot be replicated by a model lacking CO_2 forcing (e.g., Hegerl et al. 2007; Lloyd 2010, 2015). This derivational evidence, highlighting the lack of robustness of R when C is absent, is indeed relevant and bolsters climate models.

If, counterfactually, R were to remain robust even without the inclusion of core C in the model, as discussed in point b) above, this would negate any indirect confirmation (E_C) conveyed by the core to R. Thus, these studies offer another means of strengthening the robust theorem, but they achieve this by demonstrating that alternative explanations (such as natural causes like variations in solar radiation or internal climate variability) cannot alone account for the observed climate change. Interestingly, the control runs exemplify explanatory reasoning that utilises non-robustness, but their relevance for the robust theorem concerning C and R can also be evaluated by applying the two inferential rules. The inferential rules are relevant beyond robustness in this sense.

4.7 The non-monotonicity of empirical confirmation from robustness

The monotonicity of entailment was discussed in section 2.2. Here we discuss a different monotonicity concept concerning confirmation. Empirical confirmation from robustness is strictly monotonic if every new demonstration of robustness of a result confirms something, and it is weakly monotonic if demonstrating the robustness of a result never decreases the confirmation of some result. Empirical confirmation from robustness is neither because it depends on the modellers' backgound information. A given demonstration of robustness may not indirectly confirm in one set of circumstances, while con-

rather than evidence for C itself. Here, as in Lloyd (2015), the core refers to the increase in CO2 together with the other central causal relationships rather than the relationship of this increase to the temperature.

firming in another. If this is so, confirmation from robustness is not strictly monotonic. Suppose now that R_M is robust but R is not:

$$\begin{array}{ccccccccccccccc} M_i = & (\mathrm{C} & A_1 & A_2 & A_3) & & M_j & = & (\mathrm{C} & A_2 & A_4 & A_5) & \vdash & R \\ \top & | & | & | & | & & \top & & | & | & | & | & & | \\ R_M & E_C & \sim E_{A1} & \sim E_{A2} & \sim E_{A3} & & R_M & & E_C & \sim E_{A2} & \sim E_{A4} & \sim E_{A5} & & E \end{array} \tag{17}$$

Although demonstrating that R_M is robust by deriving it from M_j confirms non-empirically by increasing the conditional probability $\mathrm{p}(\mathrm{R}_M|\mathrm{C})$, it does not empirically confirm itself (unlike in case c2). While R_M is known to be closely related to $\mathrm{C\&A}_2$, R is not. To show that robustness can decrease the confirmation of a result, showing that it violates even weak monotononicity, suppose that with (17) at hand, modellers first derive

$$\begin{array}{ccccccccccccccc} M_i = & (\mathrm{C} & A_1 & A_2 & A_3) & \vdash & R & M_j & = & (\mathrm{C} & A_2 & A_4 & A_5) & \vdash & R \\ \top & | & | & | & | & & & \top & & | & | & | & | & & | \\ R_M & E_C & \sim E_{A1} & \sim E_{A2} & \sim E_{A3} & & & R_M & & E_C & \sim E_{A2} & \sim E_{A4} & \sim E_{A5} & & E \end{array} \tag{18}$$

According to DCR, this derivation increases $\mathrm{p}(\mathrm{R}|\mathrm{C})$ and $p(R|A_2)$, and thereby indirectly $\mathrm{p}(R_M|E)$. However, if they then derive

$$\begin{array}{ccccccccccccccc} M_i' = & (A_1 & A_2 & A_3) & \vdash & R & M_j & = & (\mathrm{C} & A_2 & A_4 & A_5) & \vdash & R \\ \top & | & | & | & & & \top & & | & | & | & | & & | \\ R_M & \sim E_{A1} & \sim E_{A2} & \sim E_{A3} & & & R_M & & E_C & \sim E_{A2} & \sim E_{A4} & \sim E_{A5} & & E \end{array}, \tag{19}$$

applying DDR yields $\mathrm{p}(\mathrm{C}\vdash_c\mathrm{R})=0$ and the indirect confirmation of R_M, $\mathrm{p}(R_M|E)$ decreases. The non-monotonicity highlights a key distinction between experimental/measurement robustness and derivational robustness: since indirect confirmation is neither needed nor relevant for experimental robustness, the confirmatory benefits from robustness are monotonic in that context. The non-monotonicity of R-confirmation thus arises from the fact that it is indirect.

5 Conclusion

In this paper, I have presented an account of how confirmation increases due to the robustness of results by proposing two inferential rules for reasoning with robustness. These rules dictate how researchers should adjust

their subjective conditional probabilities in response to new information regarding derivational relationships or the connections between experimental results and the components used to generate them. These rules are broadly applicable, as they closely resemble each other in experimental and modelling contexts.

However, there are important differences in how the consequences of these rules are applied across different contexts. Since experimental results are themselves empirical, applying the rules leads to an increase in empirical confirmation through a variety of evidence. In contrast, the results from models do not necessarily correspond to empirical evidence; thus, changes in conditional probabilities may strengthen the robust theorem without providing any empirical confirmation. In this sense, derivational robustness is as a non-empirical confirmation procedure. However, if empirical evidence exists for various components or results of models, then robustness can lead to empirical confirmation as well, but all relevant confirmation relations are indirect. Therefore, a major difference between experimental and derivational robustness is that the former yields empirical confirmation even without the need for indirect confirmation, while the latter does not. While the inferential rules are nearly identical in both contexts, empirical confirmation from derivational robustness requires additional resources from genuine confirmation to justify inferences concerning indirect empirical confirmation. In contrast, experimental robustness requires no such resources; a successful application of the inferential rules alone suffices for confirmation.

There are additional reasons why derivational and experimental robustness should be accounted for differently. I have argued that there are relevant differences in the roles of hypotheses and robust theorems, in the applicability of logical omniscience, and in whether the variety of evidence thesis can fail with replications.

The most important conclusion from the analysis of indirect confirmation is that if robustness shows that a result relies on the same assumptions as another empirically confirmed result, and if the robustness of either result allocates confirmation from empirical evidence based on those same assumptions, then the first result is indirectly R-confirmed through robustness.

Many critics of robustness have questioned its epistemic benefits, often tacitly assuming that proponents argue for a high absolute degree of confirmation for robust theorems. Some proponents may have contributed to this confusion by suggesting that robustness yields a high degree of absolute confirmation or that it requires a high initial degree of confirmation. How-

ever, it is clear that applying the inferential rules does not necessitate a high initial confidence in the robust theorem. In fact, the weaker the initial confidence, the greater the potential increase in confidence due to robustness. This conclusion directly follows from the inferential rules, which require that robustness demonstrate the irrelevance of at least one false auxiliary or at least one feature in an experiment.

However, these rules do not guarantee that a robust theorem or hypothesis is confirmed—either empirically or non-empirically—to a high absolute degree. My arguments for these rules, if successful, justify only an increase in confirmation. Conversely, these arguments show that critical objections to the confirmatory advantages of robustness are relevant only when aiming for a high absolute degree of confirmation. The strength of absolute confirmation ranges from very low to only slightly higher than it would be without robustness. However, since this confirmation is both incremental and indirect, it may never be very strong.

References

[1] Bogen, James, and James Woodward (1988): "Saving the Phenomena", *The Philosophical Review* 97(3): 303-352.

[2] Bovens, Luc, and Stephan Hartmann (2003): *Bayesian epistemology.* Oxford: Clarendon Press.

[3] Calcott, Brett (2011): "Wimsatt and the robustness family: Review of Wimsatt's Re-engineering Philosophy for Limited Beings", *Biology and Philosophy* 26(2): 281-293.

[4] Casini, Lorenzo, and Jürgen Landes (2024): "Confirmation by robustness analysis: A Bayesian account", *Erkenntnis* 89(1): 367-409.

[5] Claveau, François (2013): "The independence condition in the variety-of-evidence thesis", *Philosophy of Science* 80(1): 94-118.

[6] Dethier, Corey (2024): "The Unity of Robustness: Why Agreement Across Model Reports is Just as Valuable as Agreement Among Experiments", *Erkenntnis* 89(7): 2733–2752.

[7] Fitelson, Branden (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science*, 66(3), S362–S378.

[8] Forber, Patrick (2010): "Confirmation and explaining how possible", *Studies in History and Philosophy of Science Part C*, 41(1): 32-40.

[9] Fuller, Gareth, and Armin Schulz (2021) "Idealizations and Partitions: A Defense of Robustness Analysis", *European Journal for Philosophy of Science* 11(4): 1-15.

[10] Garber, Daniel (1983): "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory", in John Earman (eds.), *Testing Scientific Theories*. Minnesota: Minnesota University Press, 99-132.

[11] Gluck, Stuart (2023): "Robustness of climate models", *Philosophy of Science* 90(5): 1407-1416.

[12] Harris, Margherita (2021): "The epistemic value of independent lies: false analogies and equivocations", *Synthese* 199: 14577–14597.

[13] Harris, Margherita and Roman Frigg (2023): "Climate Models and Robustness Analysis – Part II: The Justificatory Challenge", *Handbook of the Philosophy of Climate Change*. G. Pellegrino and M. Di Paola. Netherlands, Springer International Publishing.

[14] Hegerl, Gabriele et al. (2007). Understanding and attributing climate change. In S. Solomon et al. (Eds.), *Climate change 2007: The physical science basis*. Contribution of Working Group I to the Fourth Assessment Report of the intergovernmental panel on climate change (pp. 663–745). Cambridge: Cambridge University Press

[15] Houkes, Vybo, and Krist Vaesen (2012), "Robust! - Handle with Care", *Philosophy of Science* 79: 345-364.

[16] Hudson, Robert (2013): *Seeing things: The philosophy of reliable observation*. New York: Oxford University Press.

[17] Knuuttila, Tarja, and Andrea Loettgers (2017): "Modelling as Indirect Representation? The Lotka–Volterra Model Revisited", *The British Journal for the Philosophy of Science* 68(4): 1007-1036.

[18] Kuorikoski, Jaakko, Aki Lehtinen, and Caterina Marchionni (2010), "Economic Modelling as Robustness Analysis", *British Journal for the Philosophy of Science* 61: 541-567.

[19] Landes, Jürgen (2020): "Variety of evidence and the elimination of hypotheses", *European Journal for Philosophy of Science* 10(2): 12

[20] Lehtinen, Aki (2018): "Derivational Robustness and Indirect Confirmation", *Erkenntnis* 83(3): 539-576.

[21] — (2016), "Allocating Confirmation with Derivational Robustness", *Philosophical Studies* 173: 2487-2509.

[22] Lisciandra, Chiara (2017): "Robustness analysis and tractability in modeling", *European Journal for Philosophy of Science* 7:79-95.

[23] Lloyd, Elisabeth A. (2015), "Model Robustness as a Confirmatory Virtue: The Case of Climate Science", *Studies in History and Philosophy of Science Part A* 49: 58-68.

[24] — (2010), "Confirmation and Robustness of Climate Models", *Philosophy of Science* 77: 971-984.

[25] McLoone, Brian (2025), "How to think about Indirect Confirmation", *Erkenntnis* 90: 467-481.

[26] McLoone, Brian, Orzack, Steven and Sober, Elliott (2025). "The epistemic status of derivational robustness." *European Journal for Philosophy of Science* 15(3): 1-18.

[27] Odenbaugh, Jay, and Anna Alexandrova (2011), "Buyer Beware: Robustness Analyses in Economics and Biology", *Biology and Philosophy* 26: 757-771.

[28] O'Loughlin, Ryan (2021): "Robustness reasoning in climate model comparisons", *Studies in History and Philosophy of Science Part A* 85: 34-43.

[29] O'Loughlin, Ryan, and Dan Li (2022): "Model robustness in economics: the admissibility and evaluation of tractability assumptions", *Synthese* 200(1): 1-23.

[30] Orzack, Steven, and Elliott Sober (1993): "A critical assessment of Levins's The strategy of model building in population biology (1966)", *Quarterly Review of Biology* 68(4): 533-546.

[31] Osimani, Barbara, and Jürgen Landes (2023): "Varieties of error and varieties of evidence in scientific inference", *The British Journal for the Philosophy of Science* 74(1): 117-170.

[32] Parker, Wendy S. (2011): "When climate models agree: The significance of robust model predictions", *Philosophy of Science* 78(4): 579-600.

[33] Roche, William, and Elliott Sober (2013), "Explanatoriness is Evidentially Irrelevant, Or Inference to the Best Explanation Meets Bayesian Confirmation Theory", *Analysis* 73(4): 659-668.

[34] Sayan, Erdinc (2005): "Idealizations and approximations in science, and the Bayesian theory of confirmation", in *Turkish Studies in the History and Philosophy of Science*, ed. G. Irzik and G. Güzeldere. The Netherlands: Springer, pp. 103-112.

[35] Schupbach, Jonah (2018), "Robustness Analysis as Explanatory Reasoning", *British Journal for the Philosophy of Science* 69(1): 275-300.

[36] — (2017), "Inference to the Best Explanation, Cleaned Up and made Respectable", in Kevin McCain, and Ted Poston (eds.), *Best Explanations*. Oxford: Oxford University Press, 39-61.

[37] — (2015), "Robustness, Diversity of Evidence, and Probabilistic Independence", in Uskali Mäki, Ioannis Votsis, Stéphanie Ruphy and Gerhard Schurz (eds.), *Recent Developments in the Philosophy of Science*. Heidelberg: Springer, 305-316.

[38] Schupbach, Jonah and Jan Sprenger (2011), "The Logic of Explanatory Power", *Philosophy of Science* 78(1):105-127.

[39] Schurz, Gerhard (2014), "Bayesian Pseudo-Confirmation, use-Novelty, and Genuine Confirmation", *Studies in History and Philosophy of Science* 45: 87-96.

[40] Shaffer, Michael J. (2012): *Counterfactuals and scientific realism*. Basingstoke: Palgrave Macmillan.

[41] Sprenger, Jan and Hartmann, Stephan (2019): *Bayesian Philosophy of Science*. Oxford: Oxford University Press.

[42] Stegenga, Jacob, and Tarun Menon (2017): "Robustness and Independent Evidence", *Philosophy of Science* 84(3): 414-435.

[43] Weisberg, Michael (2006), "Robustness Analysis", *Philosophy of Science* 73: 730-742.

[44] Weisberg, Michael, and Kenneth Reisman (2008): "The robust Volterra principle", *Philosophy of Science* 75(1): 106-131.

[45] Winsberg, Eric (2021): "What does robustness teach us in climate science: a re-appraisal", Synthese 198(Suppl 21): 5099–5122.

[46] Woodward, James (2006): "Some varieties of robustness", *Journal of Economic Methodology* 13(2): 219-240.